



Profiling jobseekers in Piedmont

The role of administrative and local territorial data

Monica Bosia**

Marianna Carlone*

Elisa Floredan**

Francesca Prever*

ASA Conference 2026 UniTE - Teramo

17 Settembre 2026

* Università degli studi di Torino

** Agenzia Piemonte Lavoro

Why profiling*?

- Early identification of those most at risk of long-term unemployment
- Tailoring the intensity and type of interventions to individual characteristics
- Improving the targeting of active labour policies and resource allocation

Three main approaches

Approach	Feature	Main limitation
Rule-based	Predefined rules and criteria	Limited personalisation
Caseworker-based	Caseworker assessment	Subjectivity
Statistical	Quantitative estimation of employment risk	Interpretability, transparency and dependence on data quality

*systematic use of individual information to assess re-employment prospects and tailor active labour market policies.

Profiling in Italy

In Italy, the national profiling model (ANPAL, 2022) is based on **individual characteristics** observed at the time of the DID, the formal declaration of immediate availability for work; the entry point for profiling and active labour market services.

ANPAL model variables:

- Gender
- Age
- Educational qualification
- Duration of presence in Italy
- Holding a driving licence
- Pre-DID internship experience
- Prevalent occupational status pre-DID
- Prevalent sector pre-DID
- Prevalent professional qualification pre-DID
- Number of distinct employers pre-DID
- Number of household members
- Presence of dependent children

...

FROM INDIVIDUAL PROFILING TO TERRITORIAL PROFILING

In Italy, the national profiling model (ANPAL, 2022) is based on **individual characteristics** observed at the time of the DID

Core question

How much does predictive capacity change when we add the territory profile to the person profile?

FROM THE INDIVIDUAL PROFILE...

Age
Gender
Education
Work history
Family situation
Recent experience
...



...TO THE CONTEXT IN WHICH THE PERSON LOOKS FOR WORK

Employability is an individual characteristic,
but it manifests within a local labour market.

Local Labour Market Areas

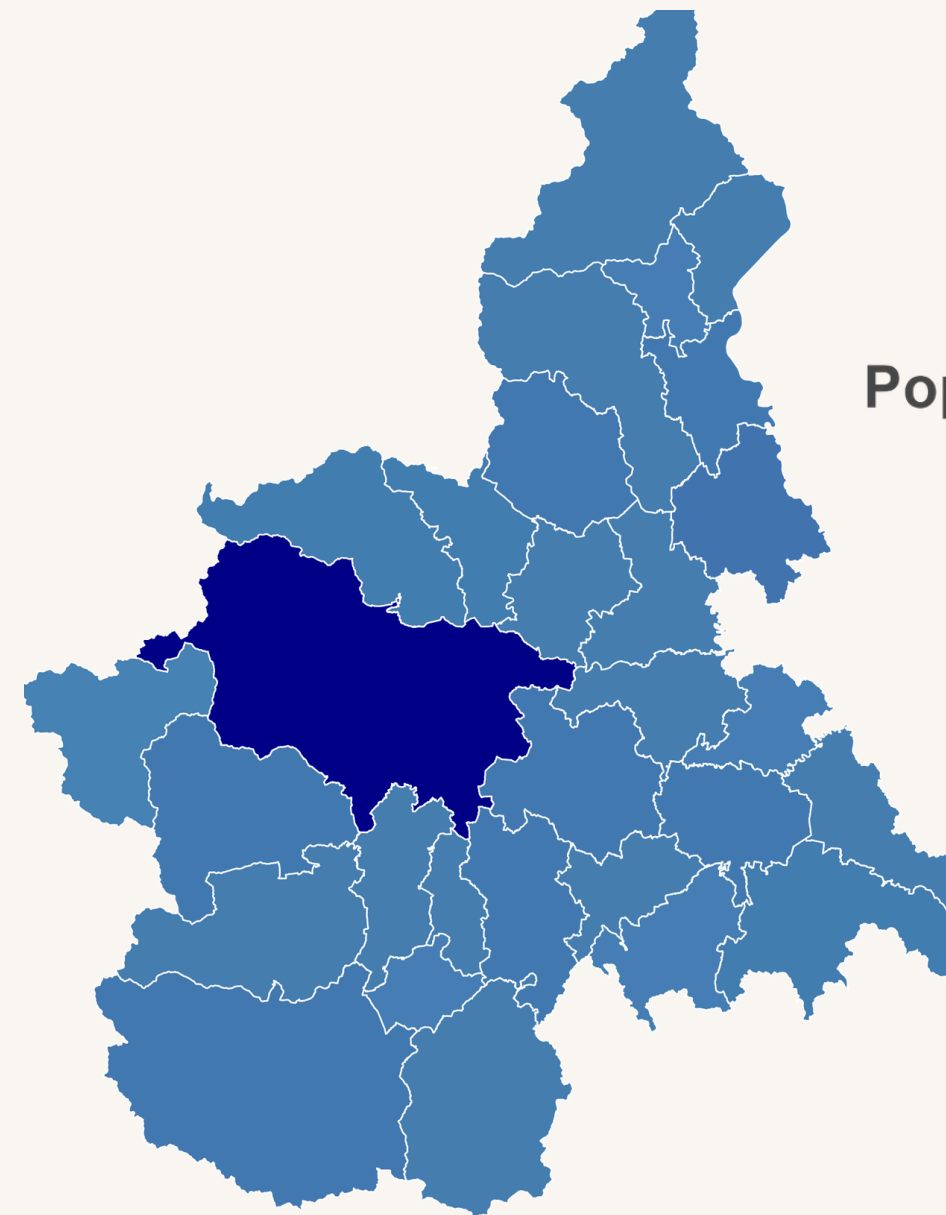


UNIVERSITÀ
DI TORINO

Territorial information is measured at the level of Local Labour Market Areas (SLL)

Local Labour Market Areas (SLL) are contiguous areas of municipalities where most people live and work, representing self-contained local labour markets.

Source: ISTAT



Popolazione dei Sistemi Locali del Lavoro

Piemonte – 2021



Why SLLs?

- Functional rather than administrative geography
- Capture the local labour market in which individuals search for and find work

Empirical strategy: three steps

1

Replicate the ANPAL model

Piedmont data - adapted specification:

- ANPAL (2022): 12-month outcome · 3,4M DID · Italy, 2018–2019
- Piedmont: 6-month outcome 163433 DID · 2022

2

Add local context

Consider a model with:

- individual characteristics
- SLL indicators

3

Increase model flexibility

Enhancing predictive performance with machine learning models:

- Random Forest
- Boosting

1

Replicate ANPAL on Piedmont data

Logistic regression: the dependent variable indicates unemployment following the DID
(1 = unemployed; 0 = employed)

$$P(Y_i = 1 | X_i) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$$

*MER = change in the predicted probability of non-reemployment for representative profiles, expressed in percentage points.

Marginal Effects at Representative values (MER)*		
Variable/Profile	Piedmont (pp)	Italy (pp)
<i>Gender and family</i>		
Woman	+1.17	+8.67
Children	-0.39	-1.20
Woman with children	+0.72	+4.50
<i>Age groups (ref. Under 24)</i>		
25-29	+1.20	+0.03
30-39	+2.64	+6.10
40-49	+6.31	+14.18
Over 50	+12.18	+20.92
<i>Recent work experience</i>		
40-49 with experience	-41.67	-9.20
Over 50 with experience	-42.54	-10.32
<i>Duration of presence in Italy</i>		
Up to 12 months	-2.59	-0.56
From 1 to 2 years	-5.61	-1.52
Over 2 years	-1.40	+1.42
Born in Italy	+14.02	-2.30

Main findings:

- **Similar patterns, different intensity:** Age and recent employment experience matter in both models.
- **Stronger role of recent work experience in Piedmont:** Recent employment is a particularly strong predictor of re-employment.
- **Territorial differences emerge already at the individual level:** The magnitude of individual characteristics varies across the two contexts.

2

Add local context

SLL Variable	Coefficient	Significance
<i>Self-employed in sole proprietorships per 100 self-employed</i>	0.0459	***
<i>Self-employed in micro-enterprises per 100 self-employed</i>	-0.1332	***
<i>Employees in high-tech and knowledge-intensive enterprises per 100 employees</i>	0.0241	***
<i>Self-employed per 100 residents (25–34 years)</i>	0.5093	***
<i>Self-employed per 100 residents (total)</i>	-0.4751	***
<i>Self-employed aged 25–34 per 100 self-employed</i>	-0.4547	***
<i>Employees aged 25–34 per 100 employees</i>	0.1766	***
<i>Employees with foreign citizenship per 100 employees (25–64 years)</i>	0.0192	**
<i>Women per 100 self-employed (25–64 years, tertiary education)</i>	0.0074	*
<i>Self-employed per 100 residents (25–64 years)</i>	-0.1023	

Model:

- Logistic regression
- Individual characteristics + 10 selected SLL indicators
- Best subset selection · AIC

Key dimensions

- Self-employment structure
- Age composition
- Foreign workers
- High-tech & knowledge-intensive employment

All associations are predictive, not causal.

3

Increase model flexibility

Method	Accuracy out-of-sample	Sensitivity	Specificity
Logistic regression — ANPAL (2022)	67.63%	55.83%	76.76%
Logistic regression	63.56%	41.63%	80.32%
Logistic regression + SLL context (Best Subset Selection)	64.14%	43.74%	79.72%
Random Forest	64.00%	45.76%	77.93%
Boosting	64.50%	47.57%	77.44%

Accuracy: share of observations correctly classified.

Sensitivity: share of non-re-employed individuals correctly identified by the model.

Specificity: share of re-employed individuals correctly identified by the model.

From logistic regression to machine learning:

- More flexible models → better identification of individuals at higher risk
- Sensitivity: 41.6% → 47.6%
- Trade-off: higher predictive flexibility · lower interpretability



UNIVERSITÀ
DI TORINO

Thank you for your attention!

Francesca Prever

ASA Conference 2026 UniTE - Teramo

17 Settembre 2026